A neurobiological pathway to smoking in adolescence: *TTC12-ANKK1-DRD2* variants and reward response

Christine Macare [a, b*], Francesca Ducci [a, b, c*], Yuning Zhang [a, b*], Barbara Ruggeri [a, b*], Tianye Jia [a, b, d], Marika Kaakinen [e, f], Gursharan Kalsi [g, h], Pimphen Charoen [i], Filippo Casoni [j, k], Jan Peters [l, m], Uli Bromberg [n], Matthew Hill [o], Jessica Buxton [p, q], Alexandra Blakemore [r], Juha Veijola [s, t], Christian Büchel [n], Tobias Banaschewski [u], Arun L.W. Bokde [v], Patricia Conrod [a, w], Herta Flor [x, y], Vincent Frouin [z], Jürgen Gallinat [aa], Hugh Garavan [ab], Penny A. Gowland [ac], Andreas Heinz [ad], Bernd Ittermann [ae], Mark Lathrop [af, ag], Jean-Luc Martinot [ah, ai], Tomáš Paus [aj], Sylvane Desrivières [a, b*], Marcus Munafò [ak*], Marjo-Riitta Järvelin [4, 5, 6, 29, 30*], Gunter Schumann [1, 2*] and the IMAGEN Consortium.

[a] Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom;
[b] Centre for Population Neuroscience and Precision Medicine (PONS), MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom;
[c] St George's, University of London, London, United Kingdom;
[d] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China and Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, China;
[e] Section of Genomics of Common Disease, Department of Medicine, Imperial College London, London, United Kingdom;
[f] Centre for Pharmacology and Therapeutics, Department of Medicine, Imperial CollegeLondon, London, United Kingdom;
[g] Clinical Physics, Barts Health NHS Trust, The Royal London Hospital, London, UK;
[h] One Small Step Gait Laboratory, Guy′s and St. Thomas' NHS Foundation Trust, Guy′s Hospital, London, UK;
[i] UCL Institute of Health Informatics, University College, London, UK;
[j] Division of Neuroscience, San Raffaele Scientific Institute, Milan 20132, Italy;
[k] Università Vita-Salute San Raffaele, Milan 20132, Italy;
[l] Intelligent Autonomous Systems Group, Department of Computer Science, Technische Universität Darmstadt, Darmstadt, Germany;
[m] Max Planck Institute for Intelligent Systems, Tübingen, Germany;
[n] University Medical Centre Hamburg-Eppendorf, House W34, 3.OG, Martinistr. 52, 20246, Hamburg, Germany;
[o] Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, United Kingdom;
[p] UCL Great Ormond Street Institute of Child Health, London;
[q] School of Life Sciences, Pharmacy and Chemistry, Kingston University, Kingston upon Thames, UK;
[r] Division of Diabetes, Endocrinology and Metabolism, Faculty of Medicine, Imperial College London, London, United Kingdom;
[s] Department of Psychiatry, Research Unit of Clinical Neuroscience, University of Oulu, Oulu, Finland;
[t] Cambridge Cognition Ltd, Cambridge, UK
[u] Department of Child and Adolescent Psychiatry, Central Institute of Mental Health, Faculty of Clinical Medicine Mannheim, University of Heidelberg, Mannheim, Germany;

[v] Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland;
[w] Department of Psychiatry, Universite de Montreal, CHU Ste Justine Hospital, Montreal, Canada;
[x] Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, Mannheim, Germany;
[y] Department of Psychology, School of Social Sciences, University of Mannheim, 68131 Mannheim, Germany;
[z] NeuroSpin, Bât 145 Gif-sur-Yvette, Paris, France;
[aa] Clinic for Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany;
[ab] Departments of Psychiatry and Psychology, University of Vermont, 05405 Burlington, Vermont, USA;
[ac] Sir Peter Mansfield Magnetic Resonance Centre, University of Nottingham, Nottingham, United Kingdom;
[ad] Clinic for Psychiatry and Psychotherapy, Campus Charité Mitte, Charité – Universitätsmedizin Berlin, Berlin, Germany;
[ae] Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany;
[af] Department of Human Genetics, McGill University, 1205 Dr Penfield Avenue, Montreal, QC, Canada;
[ag] McGill University, Genome Quebec Innovation Centre, 740 Doctor Penfield Avenue, Montreal, QC, Canada
[ah] Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 "Neuroimaging & Psychiatry", University Paris Sud – Paris Saclay, University Paris Descartes;
[ai] DIGITEO labs, Gif sur Yvette; France
[aj] Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital and Departments of Psychology and Psychiatry, University of Toronto, Toronto, Ontario, Canada;
[ak] MRC Integrative Epidemiology Unit, UK Centre for Tobacco Control Studies and School of Experimental Psychology, University of Bristol, Bristol, United Kingdom;
[al] Institute of Health Sciences, University of Oulu, Oulu, Finland;
[am] Biocenter Oulu, University of Oulu, Oulu, Finland;
[an] Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, United Kingdom;
[ao] Unit of Primary Care, Oulu University Hospital, Oulu, Finland;
[ap] Department of Children and Young People and Families, National Institute for Health and Welfare, Oulu, Finland.

*Authors had equal contribution

Corresponding Author (to whom reprint requests should be sent):
Professor Gunter Schumann, MD
Centre for Population Neuroscience and Precision Medicine (PONS), MRC Social, Genetic and Developmental Psychiatry Centre, Box 080, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF
Tel: 0044 20 7848 5314
Email: gunter.schumann@kcl.ac.uk

ABSTRACT

The *TTC12-ANKK1-DRD2* gene-cluster has been implicated in adult smoking. Here, we investigated the contribution of individual genes in the *TTC12-ANKK1-DRD2* cluster in smoking and their association with smoking-associated reward processing in adolescence. A meta-analysis of *TTC12-ANKK1-DRD2* variants and self-reported smoking behaviours was performed in four European adolescent cohorts (N=14,084). The minor G-allele of rs2236709, mapping *TTC12*, was associated with self-reported smoking (p=5.0x10^{-4}) and higher plasma cotinine levels (p=7.0x10^{-5}). This risk allele was linked to an increased ventral-striatal blood-oxygen level-dependent (BOLD) response during reward anticipation (n=1,263) and with higher *DRD2* gene expression in the striatum (p=0.013), but not with *TTC12* or *ANKK* gene expression. These data suggest a role for the *TTC12-ANKK1-DRD2* gene-cluster in adolescent smoking behaviours, provide evidence for the involvement of *DRD2* in the early stages of addiction and support the notion that genetically-driven inter-individual differences in dopaminergic transmission mediate reward sensitivity and risk to smoking.

Keywords: fMRI, genetics, IMAGEN-ALSPAC-NFBC, meta-analysis, risk taking, smoking.

INTRODUCTION

Smoking is one of the leading causes of premature death (World Health Organization, 2015). The majority of adult smokers (80-90%) initiate smoking during adolescence (Wittchen et al., 2008) and genetic factors were found to explain 44% of the individual differences (Vink et al., 2005). Adolescence is thus associated with an increased risk of developing long-lasting dependencies to nicotine and other substances (Van De Ven et al., 2010), rendering this developmental period critical for the investigation of genetic risk factors and the resulting neurobehavioral mechanisms implicated in tobacco smoking. However, our knowledge on how genes influence brain mechanisms in smoking is limited, and existing findings are often based on sample sizes not large enough to yield conclusive results (Munafo and Flint, 2011). Moreover, the investigation of smoking in adolescents is often hampered by reporting biases resulting from self-report measures (Kandel et al., 2006), thus requiring complementation by biologically-verified markers of nicotine use such as cotinine (Keskitalo et al., 2009).

Smoking in adolescence, in particular smoking initiation, is influenced by behaviours such as risk taking and impulsiveness, which reflect dissociation between the development of the subcortical reward system and a comparatively delayed maturation of cortical inhibitory functions characteristic for this developmental period (Lydon et al., 2014). In adult smokers this maturation is complete, together with the direct consequence of chronic tobacco exposure, resulting in a neurobehavioral context distinct from that in adolescents. This suggests that genetic factors and the biological processes mediating smoking behaviour may be different in adolescents and adults. For example in adults, it has previously shown that a gene-cluster containing the dopamine receptor 2 (*DRD2*), the ankyrin repeat and kinase domain containing 1 (*ANKK1*) and the tetratricopeptide repeat domain 12 (*TTC12*) genes, is associated with tobacco smoking (Ducci et al., 2011; Gelernter et al., 2006). Similar results were found in adolescents with associations stronger in adolescence than in mid-adulthood (Ducci et al.,

2011). While this observation supports the notion of shifting biological processes underlying tobacco smoking across the life span, the mechanisms by which variations in this gene-cluster exert their biological effect on smoking behaviour and smoking initiation in adolescents have not yet been elucidated.

TTC12, *ANKK1* and *DRD2* are located in close proximity on chromosome 11 in a region of high linkage disequilibrium (LD). In addition to nicotine, the gene-cluster has been implicated in alcohol and opiate addiction (Nelson et al., 2013; Xu et al., 2004; Yang et al., 2007), suggesting that its influence on smoking may be exerted through a mechanism which is not substance-specific. *TTC12* encodes for the tetratricopeptide repeat domain 12 protein, which is implicated in dopaminergic transmission and neurodevelopment (Castelo-Branco and Arenas, 2006). *ANKK1* is hypothesized to encode a signalling protein which mediates the expression of *DRD2* (Huang et al., 2009). *DRD2* in turn has a central role in regulating the dopamine reward system that mediates the reinforcing effects of all known addictive substances including nicotine mainly through striatal dopaminergic transmission (Sweitzer et al., 2012). While other dopamine-related genes have been associated to smoking behaviours (Herman et al., 2014), we were interested in further investigating the role of this genetic region in early smoking initiation (Mayhew et al., 2000) and in characterising the molecular and neurological mechanisms underlying this association.

Specifically, we aimed (i) to identify single nucleotide polymorphisms (SNPs) in the *TTC12-ANKK1-DRD2* gene-cluster, associated with self-reported cigarette smoking in adolescents through meta-analysis of smoking data from 14,084 youths and blood cotinine levels, an objective measure of nicotine exposure, in a subset of 2,540 youths, (ii) to functionally characterise relevant SNPs by measuring allele-specific gene expression in human post-mortem striatal brain tissue and (iii) to assess the genotype effect on reward sensitivity by analysing its relation with the blood-oxygen level-dependent (BOLD) response in the ventral

striatum during reward anticipation, risk taking and tobacco smoking in a subset of adolescents from the IMAGEN sample (Figure 1).

EXPERIMENTAL PROCEDURES

Participants

We included 4,512 adolescents and from the Northern Finland Birth Cohort NFBC1966 (Sabatti et al., 2009), 4,307 derived from the NFBC1986 (Vaarasmaki et al., 2009), 3674 from the Avon Longitudinal Study of Parents and Children (ALSPAC) (Golding, 1990), and 1,591 (of which 1,263 provided functional Magnetic Resonance Imaging (fMRI) and 1,085 on risk taking) from IMAGEN (Schumann et al., 2010). Demographic characteristics and phenotypic distribution of the samples are reported in Table-1.

Smoking-related phenotypes

Lifetime smoking, measuring the number and occasions of cigarette smoking adapted for adolescents who have lesser degree of exposure to smoke than adults, was self-reported in each cohort (see Supplementary-Material for details in each cohort; Ducci et al., 2011), and recoded into four categories: *Never-Tried* (never smoked), *Ever-Tried* (smoked at least once), *Smokers* (smoked more than once in the last 30 days), as previously used (Ducci et al., 2011) and *Weekly-Smokers* (smoked at least once a week in the last month). We used *Never-Tried* as our reference group to contrast with *Ever-Tried,* and subsequently validated our results by further contrasting with adolescents who smoke occasionally (*Smokers*) and regularly (*Weekly-Smokers*).

Genetic data

We chose 33 SNPs based on data availability in the cohort that was genotyped first, i.e. NFBC1966. Not all of these 33 SNPs were available in each sample. (SNPs and LD structures

for each sample can be found in Supplementary-Figure-1, MAF of the separated cohorts can be found in Supplementary Table 1).

Cotinine level

Plasma cotinine and genotypic data were available in n=2,540 ALSPAC participants (see Supplementary-Material).

Brain genotype and gene expression

We extracted gene expression levels of *TTC12*, *ANKK1* and *DRD2* in the striatum from the GEO series GSE25219 database (Kang et al., 2011). Probe clusters assessing expression across the entire transcript of *TTC12*, *ANKK1* and *DRD2* were extracted (Transcript Cluster IDs 3349453, 3349535 and 3391653, respectively). We independently replicated the association between *DRD2* expression levels (Transcript exon probe ID 3391671) and rs2236709 genotypes in 93 post-mortem cortical brain tissue samples of European ethnicity, SNPExpress (Heinzen et al., 2008).

Functional MRI

Brain activation during reward anticipation was measured using a modified monetary incentive delay (MID) fMRI task (Knutson et al., 2000) in a sample of 1,263 individuals from the IMAGEN sample. Gender, recruitment site and handedness were included as covariates. Based on previous findings (Yacubian et al., 2006), we focused on the ventral striatum (bilateral) as regions of interest (ROIs) (family-wise-error (FWE)-corrected, $p<0.05$) as 9-mm spheres from the contrast 'anticipation of large reward > anticipation of no reward' (±15, 9, -9 in Montreal Neurological Institute (MNI) space) (see Supplementary-Materials for more details).

Risk Taking

We used the Cambridge Gambling Task (CGT) from the Cambridge Cognition Neuropsychological Test Automated Battery (CANTAB; Cambridge Cognition http://www.cambridgecognition.com/).

Statistical analyses

Genetic-association analyses and meta-analysis

Genetic data analysis used Plink v1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) if not otherwise indicated. Single marker association analyses were conducted using logistic or linear regression (for binary or quantitative traits, respectively) (see also Supplementary-Materials). Each genotype was coded as the number of copies of the minor allele in all cohorts. Analyses controlled for gender, (all samples) age and site (IMAGEN).

Meta-analysis was performed using the meta-analysis procedure from Plink. Significant results after Bonferroni-correction were investigated further (see also Table-2).

Association of genotype and BOLD responses in the ventral striatum, brain activation and risk taking in the IMAGEN cohort

We carried forward the results of the confirmatory analysis to investigate associations with brain activation and behavioural phenotypes underlying reward processing. ANOVAs comparing extracted mean ventral striatum BOLD responses across genotypes for those SNPs showing associations to self-reported smoking and cotinine were carried out in SPSS version 20.0 (IBM Corp., Armonk, NY). We explored associations of the smoking risk variants within *TTC12-ANKK1-DRD2* gene-cluster with risk taking. We expected to find positive associations between genetic risk for smoking and risk taking and tested for this association one-tailed.

RESULTS

A summary of the main findings is illustrated in Figure 1.

Meta-analysis Exploring the association between TTC12-ANKKI-DRD2 and Smoking

Comparing adolescents who *Never-Tried* smoking with those who *Ever-Tried* across four cohorts, we identified a significant association with rs2236709 (Risk allele G; Odds ratio (OR) =1.10, 95% confidence interval (CI): 1.04-1.16, $p=5.0x10^{-4}$) (Figure-2 and Table-2). The minor G-allele of rs2236709 was linked to increased risk for smoking amongst subgroups of regular smokers (*Smokers* vs. *Never-Tried*: OR=1.13, 95% CI: 1.05-1.22, $p=0.001$; *Weekly-Smokers* vs. *Never-Tried*: OR=1.16, 95% CI: 1.06-1.27, $p=0.001$) (Figure-2B; Supplementary-Material Tables 2 and 3).

Association of cotinine level and self-reported smoking behavior

Higher cotinine levels correlated with a higher number of self-reported cigarettes (Spearman's $\rho=0.33$, $p=3.7x10^{-53}$), thus validating self-reported smoking. Mean cotinine levels in ng/ml (SD) for each category of smokers were: *Never-Tried*=0.73 (2.73), *Ever-Tried*=15.18 (46.51), *Smokers*=28.13 (61.64) and *Weekly-Smokers*=64.97 (82.51). One-way ANOVA revealed significant difference across lifetime smoking categories ($F(3,3319)=25.09$, $p=4.80x10^{-16}$, $R^2=0.022$). Stratification of mean plasma cotinine levels according to rs2236709 genotypes indicated a significant positive association with the minor G-allele [$F(2,2537)=9.64$, $p=6.76x10^{-5}$, $R^2=0.003$) (Figure-2C, Supplementary Table 4), thus confirming the association of rs2236709 with self-reported smoking.

Rs2236709 influences on striatal *DRD2* expression

Rs2236709 is located within *TTC12* and is in LD with the other genes in the locus, *ANKK1* and *DRD2* (see Supplementary-Material Figure-1). To investigate which of these three genes contributed to the effects of rs2236709 on smoking, we analysed their expression in post-mortem human brain samples. Figure-3A shows expression of *TTC12*, *ANKK1* and *DRD2* in the striatum across the lifetime. It was indicated that *DRD2* mRNA levels in the striatum increased steadily during the foetal stages and reached a peak in adolescence, while *ANKK1* and *TTC12* expression levels were low overall. We performed association analyses to investigate the effects of rs2236709 on striatal *DRD2*, *TTC12* and *ANKK1* expression and included developmental stage as a covariate. The minor G-allele of rs2236709 was associated with increased expression of *DRD2* in the striatum (r=.39, p=0.013, df=37; Figure-3B). Neither expression of *TTC12* nor that of *ANKK1* was influenced by rs2236709 genotypes (p=0.176 and p=0.244, respectively). There was no association of gender (p=0.627) or RNA integrity (p=0.296) with *DRD2* gene expression.

As there is, to our knowledge, no other dataset contains expression levels of *DRD2* in the striatum and rs2236709 genotype data, we independently replicated the association between *DRD2* expression levels and rs2236709 in cortical brain tissue samples of European ethnicity (Heinzen et al., 2008), and found the association to be significant ($p_{\text{uncorrected}}$=0.003, $p_{\text{corrected}}$=0.049, corrected for the number of *DRD2* probes).

Association of rs2236709 genotype and ventral striatum BOLD response during reward anticipation

As our results suggest that altered dopaminergic transmission in reward-related brain areas (e.g., the ventral striatum) may underlie the association of rs2236709 with smoking, we further measured associations of this variant with BOLD response during reward anticipation in the ventral striatum of 1,236 IMAGEN participants. Due to a skewed allele distribution of

rs2236709 (n_{AA}=706, n_{AG}=467, n_{GG}=90), we pooled heterozygotes and GG-homozygotes (G-carriers) and compared them to AA-homozygotes. Our results show a significantly higher BOLD response in the left ventral striatum of G-carriers vs. AA-homozygotes [F(1,1261)=5.803, $p_{uncorrected}$=0.016, $p_{corrected}$=0.032, R^2=0.005, Supplementary-Materials), indicating an association of the risk allele (G) with greater reward sensitivity and increased *DRD2* expression. No significant associations of rs2236709 genotypes with BOLD response in the right ventral striatum (p=0.237) were observed.

Association between Risk taking, rs2236709, and ventral striatum BOLD response during reward anticipation

As we expected a positive association between genetic risk for smoking and risk taking, this association was tested one-tailed. Risk taking positively associated with risk allele (G) of rs2236709 (r=0.056, p=0.033), and ventral striatal BOLD response (left: r=0.033, p=0.112; right: r=0.049, p=0.034). Risk taking also significantly associated with the number of occasions of lifetime cigarette smoking (r=0.056, p=0.034).

DISCUSSION

The current study carried out a neurobehavioral characterisation of the *TTC12-ANKK1-DRD2* gene-cluster and smoking in adolescence using four large datasets from different European countries. Our result demonstrated that this gene-locus exerts its effect on smoking behaviors such as smoking initiation and frequency already in the very early stages of nicotine abuse. To attain a mechanistic understanding of the associations of the minor G-allele of rs2236709 with both increased self-reported nicotine intake and higher cotinine levels in

adolescents, we found that the genetic risk factor rs2236709 regulates DRD2 gene expression, and is also associated with activation of ventral striatum during reward anticipation and risk taking, a behaviour associated with drug initiation.

DRD2 is a key molecular determinant of reward sensitivity (Sweitzer et al., 2012), which reaches its peak expression in the striatum during adolescence. High *DRD2* expression increases sensitivity to rewarding stimuli, thus increases sensitivity to the effects of substances (DiNieri et al., 2011) as well as risk taking (Cocker et al., 2012). This combination increases vulnerability for addictive behaviors (Leyton and Vezina, 2014). Thus, enhanced *DRD2* expression in adolescents is likely to be one component of risk for substance abuse and smoking in particular. Whereas *DRD2* receptor availability is reduced during the compulsive stages of substance addiction (Johnson and Kenny, 2010), early substance use has been linked to increased *DRD2* expression levels in the ventral striatum (Koob and Volkow, 2009), suggesting a differential regulation of *DRD2* in adolescents experimenting with drugs compared to established nicotine dependence in adults.

Our results suggest that carriers of rs2236709 G-risk allele are at a particularly high risk of nicotine abuse, presumably due to an allele-specific increase in *DRD2* expression, in carriers of the G-risk allele found in human post-mortem brain tissue retrieved from the striatum. While our cohorts have different ancestry the LD structure of the rs2236709 locus is similar, suggesting that the findings in the post-mortem human brains can be extended to all our cohorts. Concurrently, increased ventro-striatal response during reward anticipation was observed in carriers of the G-risk allele in functional neuroimaging analyses of 14 year-old IMAGEN participants. Together these findings suggest an enhanced (ventro-)striatal dopaminergic activity underlying the observed association. Our cognitive findings using the IMAGEN sample are in keeping with this interpretation, namely that carriers of the risk allele

showed higher risk taking behavior, which in turn was associated with increased ventral striatal activation and with increased nicotine use.

We did not detect significant association between smoking frequency and ventral striatal activity in this sample of 14 year-old adolescents in which 72% had never smoked. This might be explained by insufficient exposure to cigarettes due to young age and/or reduced availability of cigarettes compared to older cohorts, such as NFBC1966 where only 32% adolescents had never smoked at that age.

Rs2236709 is localised in the second intron of the *TTC12* gene locus and tags a haplotype that ranges from the promoter to the third exon of the *TTC12* gene. While it is correlated with *DRD2* expression in the striatum, it is not associated with the expression of *TTC12* itself. However, it is well known that even *Cis*-acting gene expression can be regulated by genetic variations that are not immediately adjacent to the gene they are regulating (Kirsten et al., 2015).

This study has some limitations: (i) different cohorts applied different measures to assess smoking behaviour, which limited us from obtaining a continuous measure consistent across studies, but instead combined information into distinct groups. (ii) Smoking habits have been subject to changing societal and legal attitudes over the last decades. Data acquisition among the different cohorts spreads over several decades, which was not accounted for. (iii) Cotinine can also account for second hand smoking, moreover the half-life of cotinine being 16-18 hr is much shorter than some of the self-reported smoking behaviours that measure longer tobacco use (Jarvis et al., 1988; Perez-Stable et al., 1995; Vartiainen et al., 2002). (iv) Genetic heterogeneity between studies should be noted as we pooled all samples of European ancestry due to limited data on adolescent smoking. Two cohorts (NFBC1966, NFBC1986) were of Finnish ancestry, and genetically distinct from the other two European cohorts (IMAGEN, ALSPAC), which might limit generalisation of our findings. Notably, despite some

differences of the minor allele frequency (MAF), the direction of the smoking-risk SNP effects is consistent across cohorts and I^2 of rs2236709 indicated no heterogeneity across cohorts.

Our work proposes a neurobiological pathway to nicotine abuse in a developmental period characterized by both vulnerability to the effects of nicotine, and great therapeutic potential to prevent or overcome its abuse (Toumbourou et al., 2007). Noting the recent success of interventions targeting behavioral risk profiles (Conrod et al., 2013), our neurobehavioral characterization provided a basis for developing interventions targeting biological mechanisms underlying nicotine abuse.

Table 1: Demographic characteristics and phenotypic distribution of the study samples.

	NFBC1966	**NFBC1986**	**ALSPAC**	**IMAGEN**		
				Whole sample	**fMRI sample**	**Risk-taking sample**
Total (N)	4512	4307	3674	1591	1263	1085
Male (in %)	47.3	47.6	51	49.6	48.9	48.9
Age in years	14	16	15	14	14	14
Country	Finland	Finland	U.K	Germany, U.K., France, Ireland	Germany, U.K., France, Ireland	Germany, U.K., France, Ireland
Year of assessment	1980	2002	2006	2008	2008	2008
Never-Tried						
n	1436	1539	1935	1139	914	771
%	31.83	35.73	52.67	71.59	72.40	71.10
Ever-Tried						
n	3076	2768	1739	452	349	314

%	68.17	64.26	47.33	28.41	27.60	28.90
Smokers						
n	757	982	891	189	146	133
%	16.78	22.8	24.25	11.88	11.40	12.30
Weekly-Smokers						
n	309	844	371	100	71	65
%	6.85	19.60	10.10	6.29	5.60	6.00

Notes Table 1: The total sample size N was decomposed into participants, who never tried smoking (*Never-Tried*) as contrasted to those who have tried smoking at least once (*Ever-Tried*); this comparison is mutually exclusive. Out of those *Ever-Tried* participants, smokers and weekly smokers were taken; percentages refer to the total sample size (N).

Table 2: Association between SNPs spanning the *TTC12-ANKK1-DRD2* gene-cluster and smoking for *Ever-Tried* vs *Never-Tried*.

BP	**Gene**	**SNP**	**Location**	**MAF**	**Ref allele/ Other allele**	**N**		***P***		***OR***		***Q***	***I^2***
						Samples	**Individuals**	**(FEM)**	**(REM)**	**(FEM)**	**(REM)**		
112685412	*TTC12*	rs4517559	flanking 5'UTR	0.3736	C/T	4	13553	**0.002**		1.08		0.27	23.84
112691986	*TTC12*	rs2236709	Intron	0.2588	G/A	4	14081	**0.0005***		1.1		0.66	0
112693407	*TTC12*	rs7927508	Intron	0.3667	G/A	3	9762	**0.009**		1.08		0.15	46.93
112694133	*TTC12*	rs2156486	Intron	0.1814	G/T	4	13555	**0.008**		0.92		0.83	0
112699378	*TTC12*	rs723077	Coding, Met73Leu	0.4894	C/A	3	9771	0.06		0.94		0.63	0
112704356	*TTC12*	rs10502172	Intron	0.4596	T/C	4	13675	**0.004**		0.93		0.16	41.71

112705919	*TTC12*	rs2303380	Intron splice site	0.3819	G/A	3	9771	0.49		1.02		0.18	41.76
112716539	*TTC12*	rs2288159	Intron	0.1561	T/G	3	9768	0.13		1.06		0.63	0
112721133	*TTC12*	rs4987094	Intron	0.1004	A/G	3	9777	0.12		1.07		0.67	0
112735810	*TTC12*	rs2276070	Intron	0.1541	T/C	3	9774	0.13		1.06		0.65	0
112739889	*TTC12*	rs719802	Intron	0.3953	T/C	3	9772	0.36		1.03		0.33	8.76
112739985	*TTC12*	rs719804	Intron	0.249	G/A	3	9759	0.06		1.07		0.76	0
112749387	*TTC12*	rs2282511	flanking 3'UTR	0.3455	A/C	3	11764		0.11		1.08	0.09	57.82
112754346	*TTC12*	rs754672	flanking 3'UTR	0.4894	T/C	4	13552	**0.002**		0.92		0.48	0

112761718	*ANKK1*	rs877138	flanking 5'UTR	0.3431	G/A	3	9777	0.25		1.04		0.37	0
112768580	*ANKK1*	rs4590907	Intron	0.1436	G/T	4	13577	0.23		1.04		0.76	0
112772031	*ANKK1*	rs7118900	Coding, Ala239Thr	0.1896	A/G	3	9767	0.24		0.95		0.88	0
112775225	*ANKK1*	rs4938016	Coding, Gly442Arg	0.4452	G/C	3	11966	**0.002**		1.09		0.5	0
112775370	*ANKK1*	rs2734849	Coding, His490Arg	0.4886	G/A	4	13580	**0.01**		0.94		0.3	18.9
112776038	*ANKK1*	rs1800497	Coding, Glu713Lys, TaqIA	0.2026	A/G	4	13549	0.47		0.98		0.71	0
112788669	*DRD2*	rs6277	Coding, Pro219Pro	0.12201	A/G	3	11981	**0.004**		0.92		0.57	0

112801119	*DRD2*	rs1076563	Intron	0.4127	C/A	3	9773	0.07		0.95		0.78	0
112803549	*DRD2*	rs2471857	Intron	0.1577	T/C	3	9772	0.38		0.96		0.7	0
112815891	*DRD2*	rs7125415	Intron	0.09914	T/C	3	9775	0.09		1.08		0.43	0
112818599	*DRD2*	rs4648318	Intron	0.2538	C/T	3	9768	**0.02**		1.08		0.53	0
112824662	*DRD2*	rs4274224	Intron	0.4823	G/A	3	9771	0.48		1.02		0.42	0
112829684	*DRD2*	rs4581480	Intron	0.09919	C/T	3	9768	0.42		1.04		0.28	20.71
112834984	*DRD2*	rs7131056	Intron	0.4205	C/A	3	9762	0.14		1.05		0.37	0
112846601	*DRD2*	rs4938019	Intron	0.1447	C/T	4	13588	0.29		1.04		0.41	0

112852165	*DRD2*	rs1236428 3	flanking 5'UTR	0.07563	G/A	3	9765	0.94		0.1		0.98	0
112857971	*DRD2*	rs1089155 6	intergenic	0.1758	T/G	3	9772	0.51		1.03		0.35	5.8
112860946	*DRD2*	rs6589377	intergenic	0.3776	G/A	4	13581	0.66		1.01		0.34	10.46
112863421	*DRD2*	rs4482060	intergenic	0.30032	T/A	3	11955	0.39		1.03		0.48	0

Notes Table 2: Association between SNPs spanning the *TTC12-ANKK1-DRD2* gene-cluster and self-reported smoking behaviour. *Ever-Tried* (N=8722) are compared to *Never-Tried* (N=6049) (binary logistic regression). Results from NFBC1966, ALSPAC, NFBC1986 and IMAGEN have been combined using meta-analysis. FEM=Fixed effect model, REM=Random effect model, *Q*=Cochrane's *Q* statistic, I^2=heterogeneity index (0-100), N=number of study samples. Significant p values are in bold, of these those that remain significant after Bonferroni correction for multiple testing are indicated by an asterisk. Thick lines indicate boundary between Linkage Disequilibrium blocks.

Figure legends:

Figure 1: Study aims, statistical analysis strategy, and results.

Figure 2: Association between 33 single nucleotide polymorphisms (SNPs) covering the chromosome 11q23 *TTC12-ANKK1-DRD2* gene-cluster and self-reported smoking behavior in adolescence. Results of the meta-analyses across the NFBC1966, NFBC1986, ALSPAC and IMAGEN cohorts are reported. *p*-values were meta-*p*-values computed under fixed effect models. Given the multiple testing correction for all SNPs a significance threshold of .0015 was used. The blue diamond indicates the most significantly associated SNP. This was rs2236709 with $p=5.0x10^{-4}$ for *Ever-Tried* vs. *Never-Tried*, rs2236709 with *p*=.001 for *Smokers* vs. *Never-Tried* and rs2236709 with *p*=.001 for *Weekly Smokers* vs. *Never-Tried*. For other SNPs, diamonds are colored in a white-to-red scale corresponding to R^2 values from 0 to 1 with the most significant SNPs. The SNP position refers to National Center for Biotechnology Information build 35. Estimated recombination rates are from HapMap and gene annotations are from UCSC genome browser with build. **(B)** Forest plots for rs2236709 for the comparisons of the reference group (i.e. *Never-Tried*) with *Ever-Tried*, *Smokers* and *Weekly-Smokers*. Box areas are proportional to the weight of the individual study. The overall summary odds ratios (OR) computed with a fixed effect model are represented by a diamond whose width indicate the 95% confidence interval (CI). Results of the analyses were as follows: *Ever-Tried* vs. *Never-Tried* OR=1.10, 95%CI 1.04-1.16, $p=.5.0x10^{-4}$; *Smokers* vs. *Never-Tried* OR=1.13, 95% CI 1.05-1.22, *p*=.001 and *Weekly-Smokers* vs. *Never-Tried* OR=1.16, 95% CI 1.06-1.27, *p*=.001 on the left, middle and right forest plots, respectively. **(C, left plot)** Association between 33 SNPs and cotinine level in the ALSPAC cohort. Blue diamond indicates the most significantly associated SNP (rs2236709), as determined by linear regression analyses. Other diamonds are color-coded in a white-to-orange scale corresponding to increasing R^2 values of

the respective SNP with rs2236709. The SNP position refers to National Center for Biotechnology Information build 35. Estimated recombination rates are from HapMap and gene annotations are from UCSC genome browser with build 35 coordinates. **(C, right plot)** Mean cotinine level (SE) according to rs2236709 genotypes (linear regression analysis: β=0.11, p=.5.0x10^{-4}). $Mean_{AA}$=6.35, SE=.72, n=1489; $Mean_{AG}$=7.94, SE=1.18, n=900; and $Mean_{GG}$=19.62, SE=4.86, n=151; post-hoc comparisons: AA vs AG, p=.843, R^2=.0005, AA vs GG, p=3.67E-5, R^2=.02 and AG vs GG, p=.001, R^2=.013.

Figure 3: (A) Expression of *DRD2*, *TTC12* and *ANKK1* in the human striatum across the lifetime. The x-axis represents different stages of fetal and postnatal development, as indicated. PCW=post-conception weeks, M=months, Y=years. **(B)** Striatal expression of *DRD2* corrected for effects of developmental stage and stratified by rs2236709 genotypes. The dotted line visualizes the linear regression against the number of minor G-alleles (n_{AA}=22, n_{AG}=16 and, n_{GG}=5).

REFERENCES

Castelo-Branco, G., Arenas, E., 2006. Function of Wnts in dopaminergic neuron development. Neuro-degenerative diseases 3, 5.

Cocker, P.J., Dinelle, K., Kornelson, R., Sossi, V., Winstanley, C.A., 2012. Irrational choice under uncertainty correlates with lower striatal D2/3 receptor binding in rats. The Journal of Neuroscience 32, 15450-15457.

Conrod, P.J., O'Leary-Barrett, M., Newton, N., et al., 2013. Effectiveness of a selective, personality-targeted prevention program for adolescent alcohol use and misuse: A cluster randomized controlled trial. JAMA psychiatry 70, 334-342.

DiNieri, J.A., Wang, X., Szutorisz, H., Spano, S.M., Kaur, J., Casaccia, P., Dow-Edwards, D., Hurd, Y.L., 2011. Maternal cannabis use alters ventral striatal dopamine D2 gene regulation in the offspring. Biological Psychiatry 70, 763-769.

Ducci, F., Kaakinen, M., Pouta, A., Hartikainen, A.L., Veijola, J., Isohanni, M., Charoen, P., Coin, L., Hoggart, C., Ekelund, J., Peltonen, L., Freimer, N., Elliott, P., Schumann, G., Jarvelin, M.R., 2011. TTC12-ANKK1-DRD2 and CHRNA5-CHRNA3-CHRNB4 influence different pathways leading to smoking behavior from adolescence to mid-adulthood. Biological Psychiatry 69, 650-660.

Gelernter, J., Yu, Y., Weiss, R., Brady, K., Panhuysen, C., Yang, B.Z., Kranzler, H.R., Farrer, L., 2006. Haplotype spanning TTC12 and ANKK1, flanked by the DRD2 and NCAM1 loci, is strongly associated to nicotine dependence in two distinct American populations. Human Molecular Genetics 15, 3498-3507.

Golding, J., 1990. Children of the nineties. A longitudinal study of pregnancy and childhood based on the population of Avon (ALSPAC). West Engl Med J 105, 80-82.

Heinzen, E.L., Ge, D., Cronin, K.D., Maia, J.M., Shianna, K.V., Gabriel, W.N., Welsh-Bohmer, K.A., Hulette, C.M., Denny, T.N., Goldstein, D.B., 2008. Tissue-specific genetic control of splicing: implications for the study of complex traits. PLoS Biol 6, e1.

Herman, A.I., DeVito, E.E., Jensen, K.P., Sofuoglu, M., 2014. Pharmacogenetics of nicotine addiction: role of dopamine. Pharmacogenomics 15, 221-234.

Huang, W., Payne, T.J., Ma, J.Z., Beuten, J., Dupont, R.T., Inohara, N., Li, M.D., 2009. Significant association of ANKK1 and detection of a functional polymorphism with nicotine dependence in an African-American sample. Neuropsychopharmacology 34, 319-330.

Jarvis, M.J., Russell, M.A., Benowitz, N.L., Feyerabend, C., 1988. Elimination of cotinine from body fluids: implications for noninvasive measurement of tobacco smoke exposure. American journal of public health 78, 696-698.

Johnson, P.M., Kenny, P.J., 2010. Dopamine D2 receptors in addiction-like reward dysfunction and compulsive eating in obese rats. Nature neuroscience 13, 635-641.

Kandel, D.B., Schaffran, C., Griesler, P.C., Hu, M.-C., Davies, M., Benowitz, N., 2006. Salivary cotinine concentration versus self-reported cigarette smoking: three patterns of inconsistency in adolescence. Nicotine & tobacco research 8, 525-537.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M.B., Krsnik, Z., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S.N., Vortmeyer, A., Weinberger, D.R., Mane, S., Hyde, T.M., Huttner, A., Reimers, M., Kleinman, J.E., Sestan, N., 2011. Spatio-temporal transcriptome of the human brain. Nature 478, 483-489.

Keskitalo, K., Broms, U., Heliovaara, M., Ripatti, S., Surakka, I., Perola, M., Pitkaniemi, J., Peltonen, L., Aromaa, A., Kaprio, J., 2009. Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNB4) on chromosome 15. Hum Mol Genet 18, 4007-4012.

Kirsten, H., Al-Hasani, H., Holdt, L., Gross, A., Beutner, F., Krohn, K., Horn, K., Ahnert, P., Burkhardt, R., Reiche, K., 2015. Dissecting the Genetics of the Human Transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci. Human molecular genetics.

Knutson, B., Westdorp, A., Kaiser, E., Hommer, D., 2000. FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage 12, 20-27.

Koob, G.F., Volkow, N.D., 2009. Neurocircuitry of addiction. Neuropsychopharmacology 35, 217-238.

Leyton, M., Vezina, P., 2014. Dopamine ups and downs in vulnerability to addictions: a neurodevelopmental model. Trends in pharmacological sciences 35, 268-276.

Lydon, D.M., Wilson, S.J., Child, A., Geier, C.F., 2014. Adolescent brain maturation and smoking: What we know and where we're headed. Neuroscience & Biobehavioral Reviews 45, 323-342.

Mayhew, K.P., Flay, B.R., Mott, J.A., 2000. Stages in the development of adolescent smoking. Drug and alcohol dependence 59 Suppl 1, S61-81.

Munafo, M.R., Flint, J., 2011. Dissecting the genetic architecture of human personality. Trends in cognitive sciences 15, 395-400.

Nelson, E.C., Lynskey, M.T., Heath, A.C., Wray, N., Agrawal, A., Shand, F.L., Henders, A.K., Wallace, L., Todorov, A.A., Schrage, A.J., Saccone, N.L., Madden, P.A., Degenhardt, L., Martin, N.G., Montgomery, G.W., 2013. ANKK1, TTC12, and NCAM1 Polymorphisms and Heroin Dependence: Importance of Considering Drug Exposure. JAMA psychiatry, 1-9.

World Health Organisation, 2015. WHO report on the global tobacco epidemic 2015. World Health Organization.

Perez-Stable, E.J., Benowitz, N.L., Marin, G., 1995. Is serum cotinine a better measure of cigarette smoking than self-report? Preventive medicine 24, 171-179.

Sabatti, C., Service, S.K., Hartikainen, A.L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M., Sovio, U., Ruokonen, A., Laitinen, J., Jakkula, E., Coin, L., Hoggart, C., Collins, A., Turunen, H., Gabriel, S., Elliot, P., McCarthy, M.I., Daly, M.J., Jarvelin, M.R., Freimer, N.B., Peltonen, L., 2009. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nature genetics 41, 35-46.

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., Conrod, P., Dalley, J., Flor, H., Gallinat, J., 2010. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Molecular Psychiatry 15, 1128-1139.

Sweitzer, M.M., Donny, E.C., Hariri, A.R., 2012. Imaging genetics and the neurobiological basis of individual differences in vulnerability to addiction. Drug and Alcohol Dependence 123 Suppl 1, S59-71.

Toumbourou, J.W., Stockwell, T., Neighbors, C., Marlatt, G.A., Sturge, J., Rehm, J., 2007. Adolescent health 4 - Interventions to reduce harm associated with adolescent substance use. Lancet 369, 1391-1401.

Vaarasmaki, M., Pouta, A., Elliot, P., Tapanainen, P., Sovio, U., Ruokonen, A., Hartikainen, A.L., McCarthy, M., Jarvelin, M.R., 2009. Adolescent manifestations of metabolic syndrome among children born to women with gestational diabetes in a general-population birth cohort. American journal of epidemiology 169, 1209-1215.

Van De Ven, M.O., Greenwood, P.A., Engels, R.C., Olsson, C.A., Patton, G.C., 2010. Patterns of adolescent smoking and later nicotine dependence in young adults: a 10-year prospective study. Public Health 124, 65-70.

Vartiainen, E., Seppala, T., Lillsunde, P., Puska, P., 2002. Validation of self reported smoking by serum cotinine measurement in a community-based study. Journal of epidemiology and community health 56, 167-170.

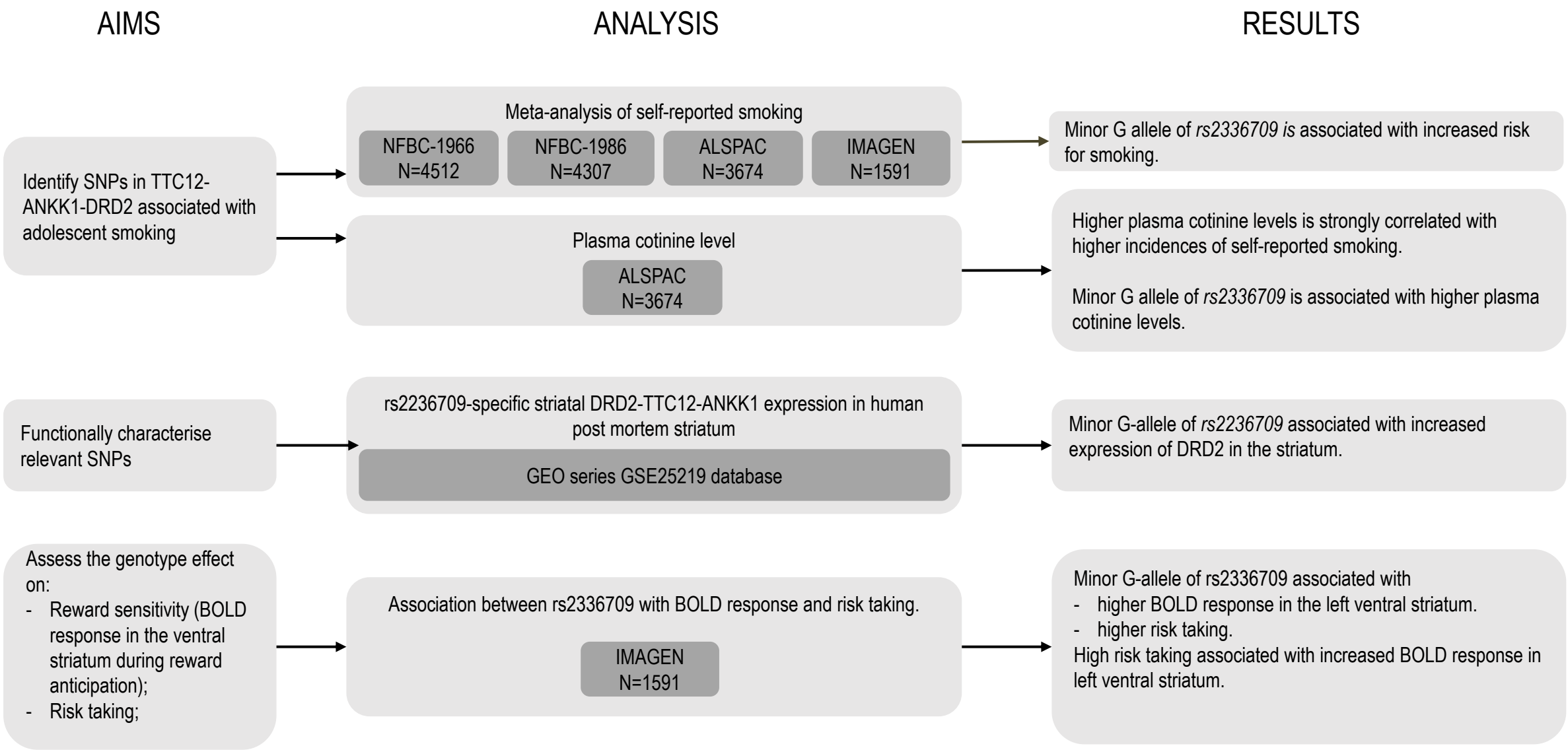
Vink, J.M., Willemsen, G., Boomsma, D.I., 2005. Heritability of smoking initiation and nicotine dependence. Behavior Genetics 35, 397-406.

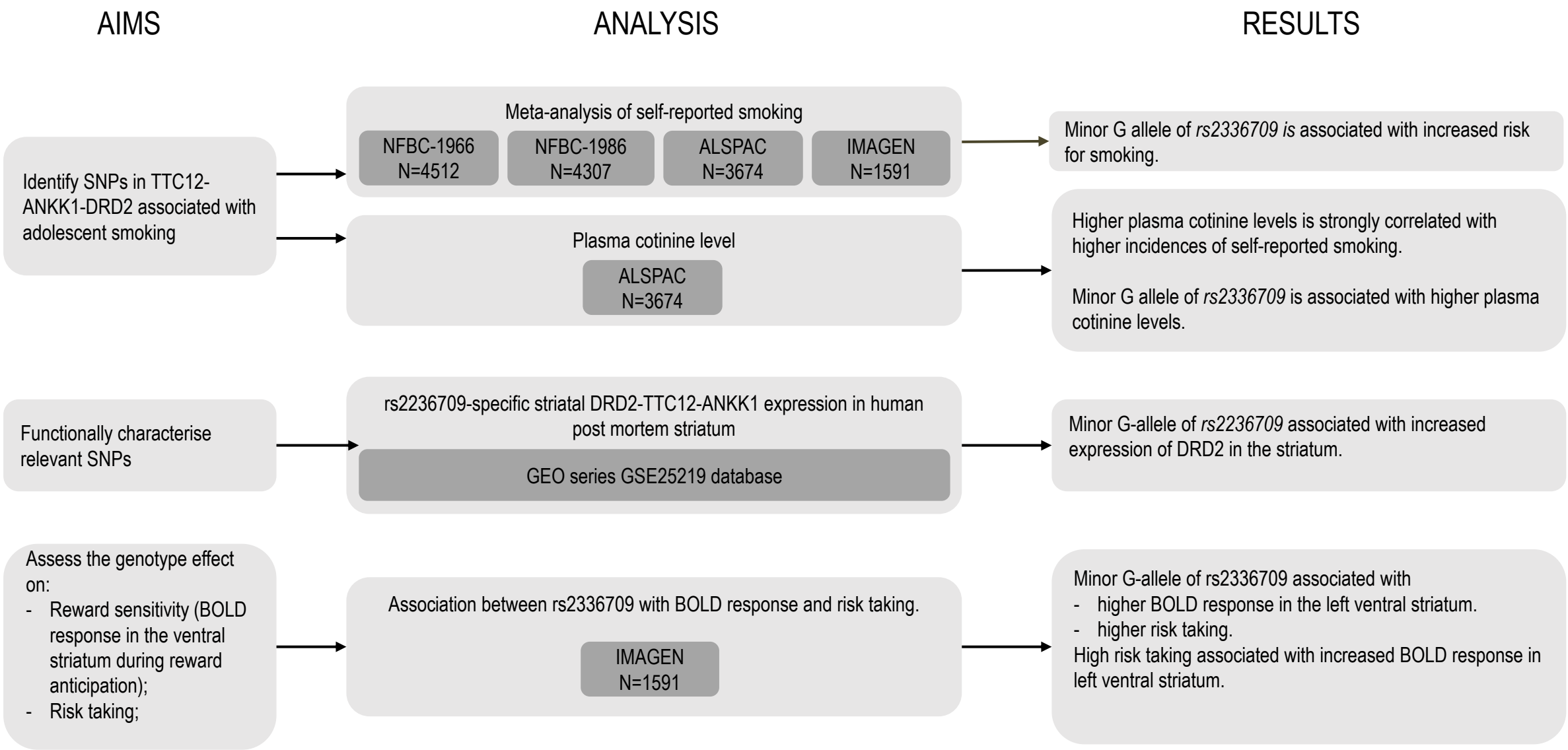
Wittchen, H.U., Behrendt, S., Hofler, M., Perkonigg, A., Lieb, R., Buhringer, G., Beesdo, K., 2008. What are the high risk periods for incident substance use and transitions to abuse and dependence? Implications for early intervention and prevention. Int J Methods Psychiatr Res 17 Suppl 1, S16-29.

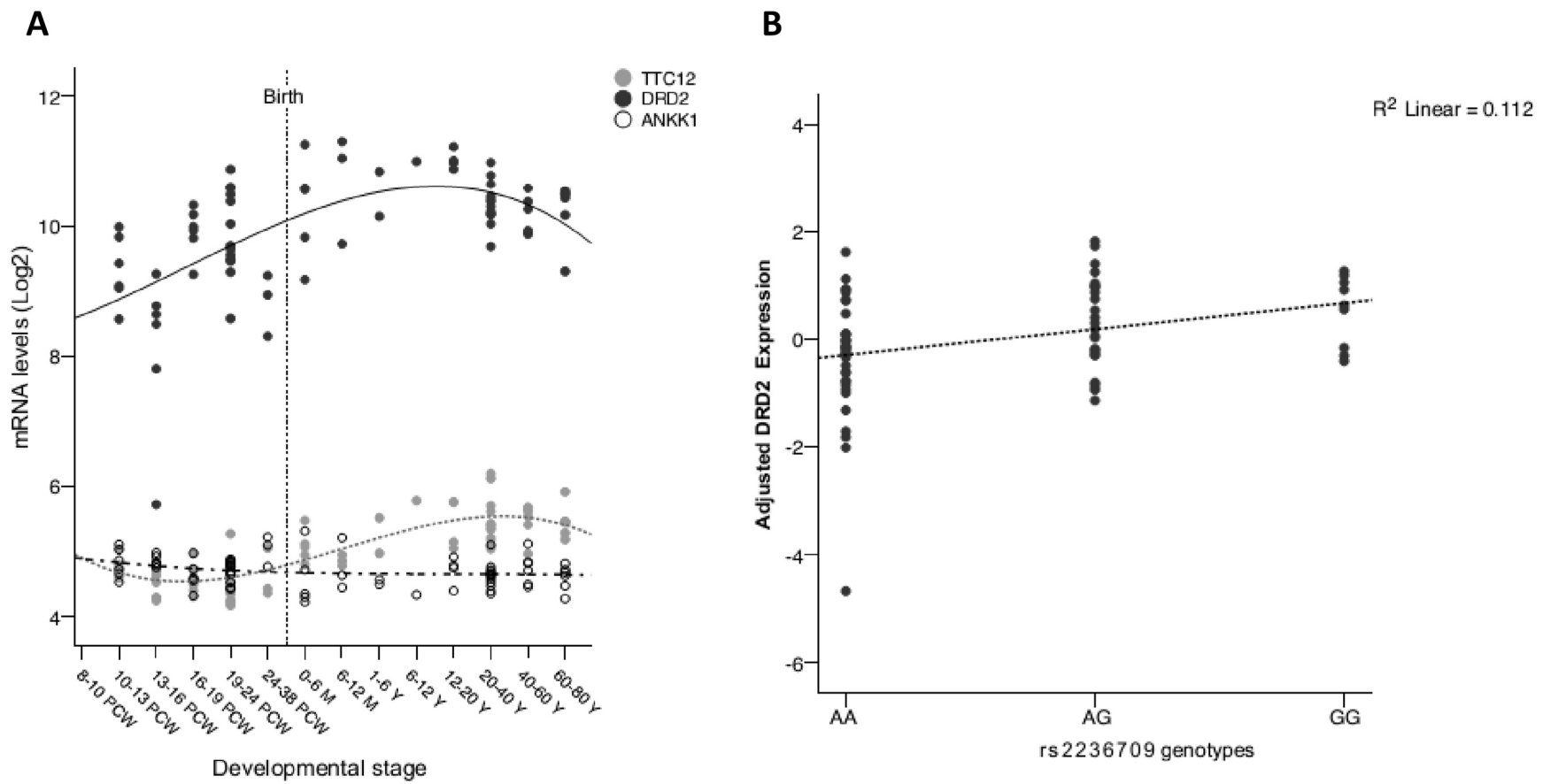
Xu, K., Lichtermann, D., Lipsky, R.H., Franke, P., Liu, X., Hu, Y., Cao, L., Schwab, S.G., Wildenauer, D.B., Bau, C.H., Ferro, E., Astor, W., Finch, T., Terry, J., Taubman, J., Maier, W., Goldman, D., 2004. Association of specific haplotypes of D2 dopamine receptor gene with vulnerability to heroin dependence in 2 distinct populations. Archives of General Psychiatry 61, 597-606.

Yacubian, J., Glaescher, J., Schroeder, K., Sommer, T., Braus, D.F., Buechel, C., 2006. Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. Journal of Neuroscience 26, 9530-9537.

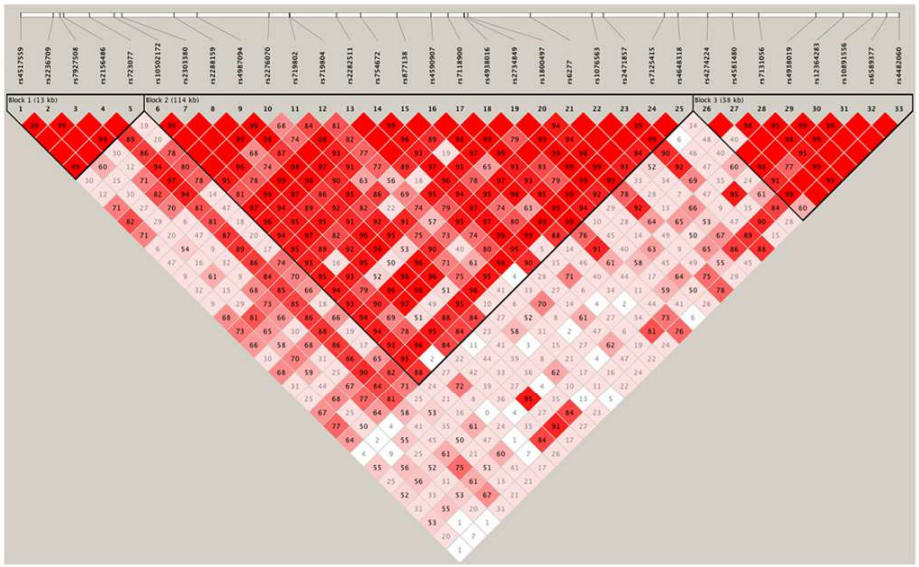
Yang, B.Z., Kranzler, H.R., Zhao, H., Gruen, J.R., Luo, X., Gelernter, J., 2007. Association of haplotypic variants in DRD2, ANKK1, TTC12 and NCAM1 to alcohol dependence in independent case control and family samples. Hum Mol Genet 16, 2844-2853.


AIMS
ANALYSIS
RESULTS
Identify SNPs in TTC12-ANKK1-DRD2 associated with adolescent smoking
Meta-analysis of self-reported smoking
NFBC-1966 N=4512
NFBC-1986 N=4307
ALSPAC N=3674
IMAGEN N=1591
Minor G allele of rs2336709 is associated with increased risk for smoking.
Plasma cotinine level
ALSPAC N=3674
Higher plasma cotinine levels is strongly correlated with higher incidences of self-reported smoking.
Minor G allele of rs2336709 is associated with higher plasma cotinine levels.
Functionally characterise relevant SNPs
rs2236709-specific striatal DRD2-TTC12-ANKK1 expression in human post mortem striatum
GEO series GSE25219 database
Minor G-allele of rs2236709 associated with increased expression of DRD2 in the striatum.
Assess the genotype effect on:
- Reward sensitivity (BOLD response in the ventral striatum during reward anticipation);
- Risk taking;
Association between rs2336709 with BOLD response and risk taking.
IMAGEN N=1591
Minor G-allele of rs2336709 associated with
- higher BOLD response in the left ventral striatum.
- higher risk taking.
High risk taking associated with increased BOLD response in left ventral striatum.


AIMS
ANALYSIS
RESULTS
Identify SNPs in TTC12-ANKK1-DRD2 associated with adolescent smoking
Meta-analysis of self-reported smoking
NFBC-1966 N=4512
NFBC-1986 N=4307
ALSPAC N=3674
IMAGEN N=1591
Minor G allele of rs2336709 is associated with increased risk for smoking.
Plasma cotinine level
ALSPAC N=3674
Higher plasma cotinine levels is strongly correlated with higher incidences of self-reported smoking.
Minor G allele of rs2336709 is associated with higher plasma cotinine levels.
Functionally characterise relevant SNPs
rs2236709-specific striatal DRD2-TTC12-ANKK1 expression in human post mortem striatum
GEO series GSE25219 database
Minor G-allele of rs2236709 associated with increased expression of DRD2 in the striatum.
Assess the genotype effect on:
- Reward sensitivity (BOLD response in the ventral striatum during reward anticipation);
- Risk taking;
Association between rs2336709 with BOLD response and risk taking.
IMAGEN N=1591
Minor G-allele of rs2336709 associated with
- higher BOLD response in the left ventral striatum.
- higher risk taking.
High risk taking associated with increased BOLD response in left ventral striatum.

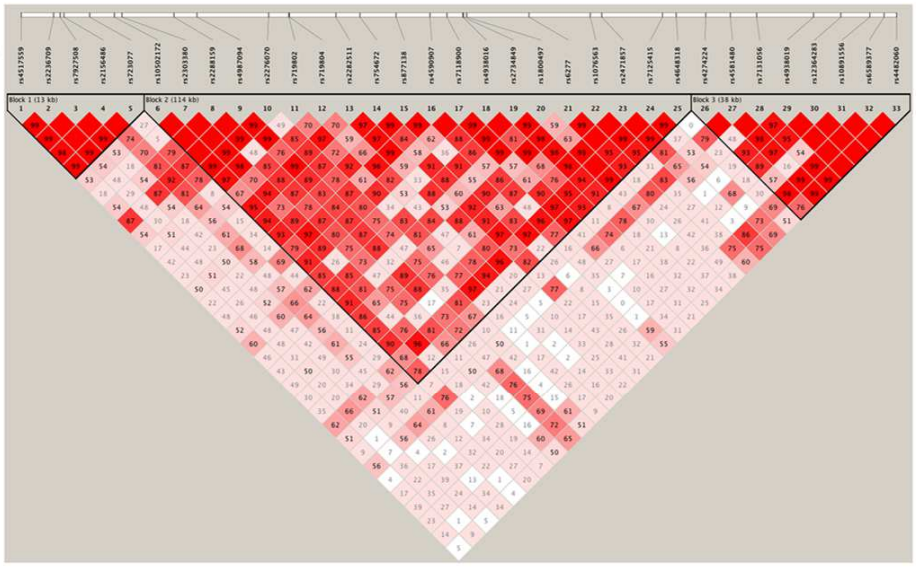

A
B
TTC12
DRD2
ANKK1
Birth
mRNA levels (Log2)
12
10
8
6
4
8-10 PCW
10-13 PCW
13-16 PCW
16-19 PCW
19-24 PCW
24-38 PCW
0-6 M
6-12 M
1-6 Y
6-12 Y
12-20 Y
20-40 Y
40-60 Y
60-80 Y
Developmental stage
R2 Linear = 0.112
Adjusted DRD2 Expression
4
2
0
-2
-4
-6
AA
AG
GG
rs2236709 genotypes

A



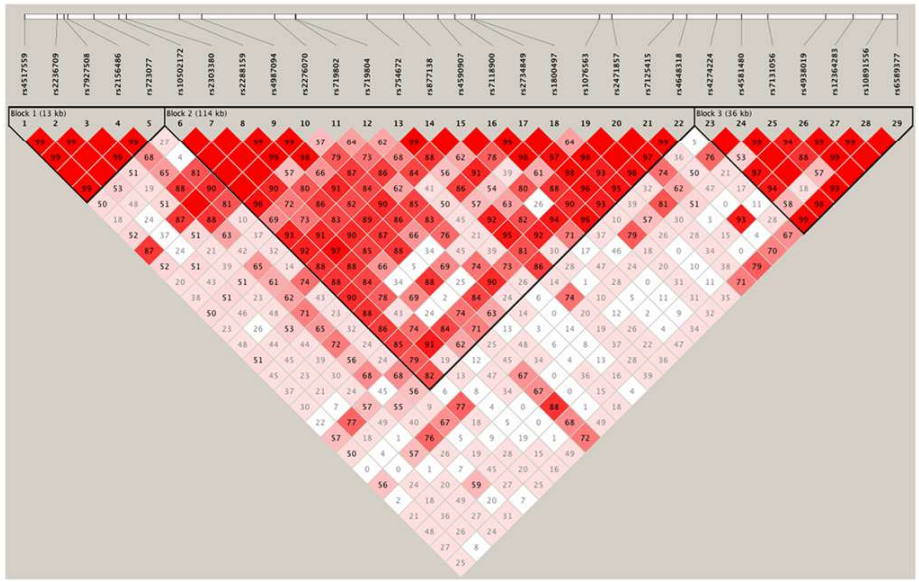
NFBC1966

B



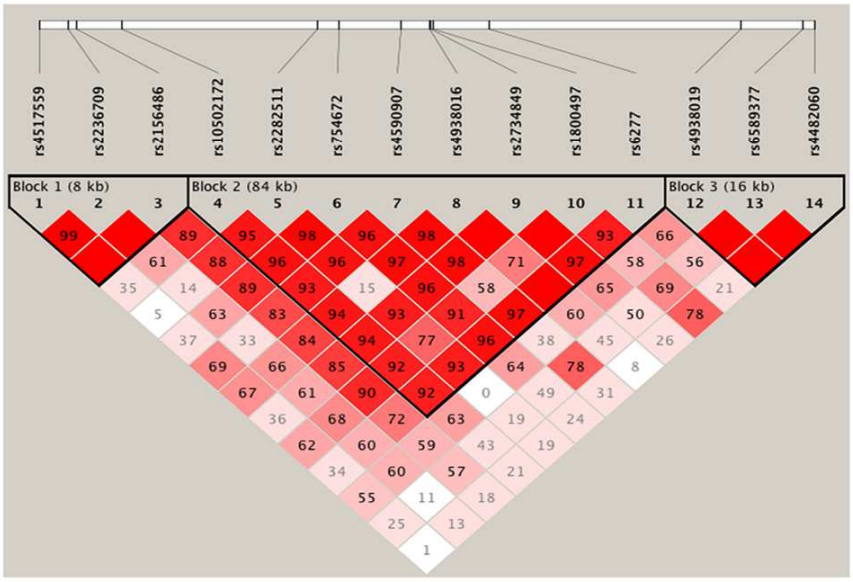
ALSPAC

C



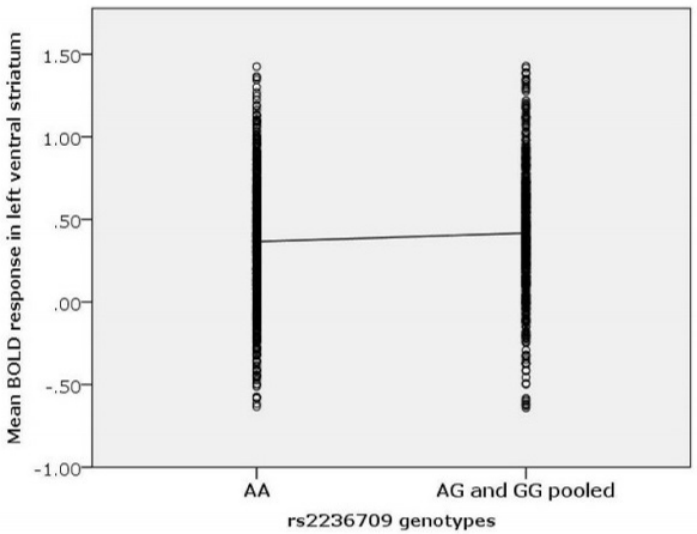
IMAGEN

D

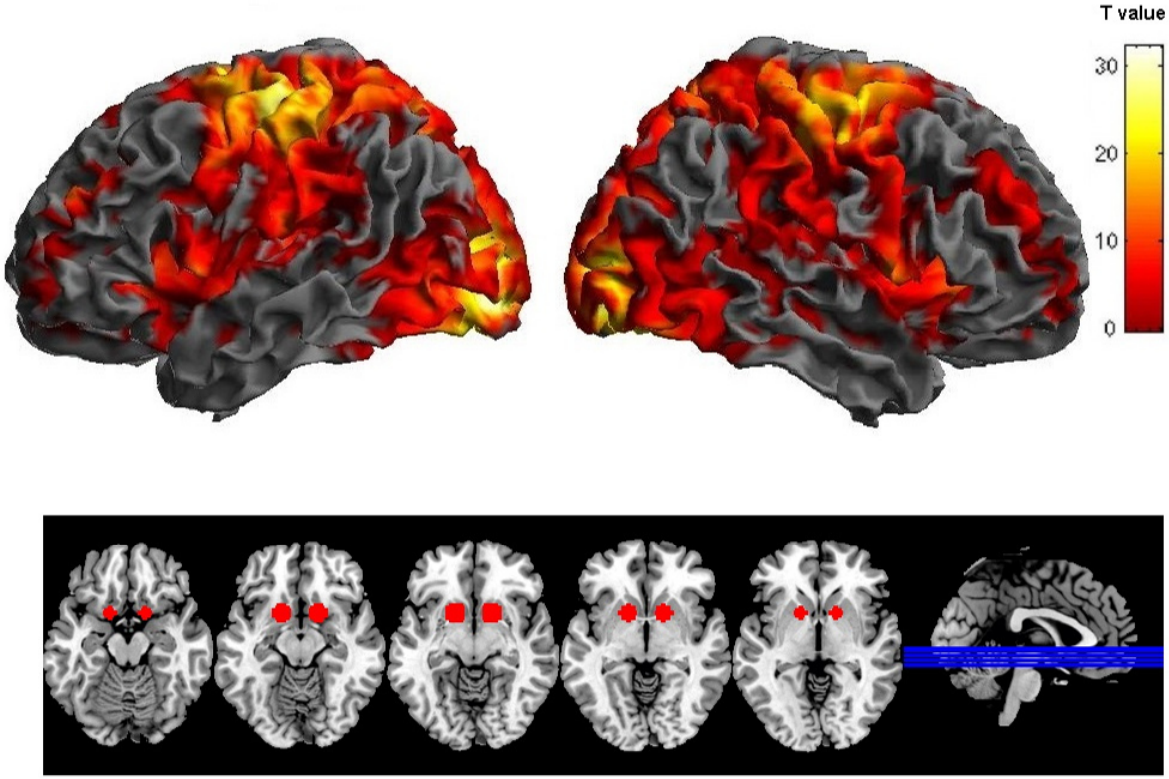


NFBC1986

A


Mean BOLD response in left ventral striatum
1.50
1.00
.50
.00
-.50
-1.00
AA
AG and GG pooled
rs2236709 genotypes

B


T value
30
20
10
0

Supplementary Material

METHODS AND MATERIALS

For all studies, informed consent had been obtained and ethical approval for the study was granted by the local Research Ethics committees.

Assessment of self-reported smoking and definition of smoking categories in the four samples

NFBC1966: At age 14, participants reported if they have ever smoked and how much via a postal questionnaire using eight options: 1) 'Never'; 2) 'I have tried once'; 3) 'I have tried twice or more'; 4) 'I smoke occasionally'; 5) 'I smoke about twice a week'; 6) 'I smoke between 1–5 cigarettes a day'; 7) 'I smoke between 6–10 cigarettes a day' and 8) 'I smoke more than 10 cigarettes a day'.

The category *Ever-Tried* included groups 2-8; *Smokers* included groups 4-8; *Weekly-Smokers* included groups 5-8, and *Never-Tried* included group 1.

NFBC1986: At age 16, participants were asked by postal questionnaires: 'Have you ever smoked or used snuff in your life?' with four possible answers: 1) 'no',2) 'yes but only tried',3) 'yes, smoke' and4) 'yes, snuff'.

Participants in group 1 were classified as *Never-Tried*; those in groups 2-3 were classified as *Ever-Tried and* those in group 3 as *Smokers*. Participants in group 4 (N=36) were excluded. Cohort members were also asked how often they smoked and participants who were smoking every week or more were classified as *Weekly-Smokers*.

ALSPAC: At age 15, participants were invited to attend a face-to-face interview about tobacco use. Participants were asked multiple questions including: 'Have you ever tried a cigarette, even a puff?'. Participants who responded affirmatively to this question were classified as *Ever-Tried* and participants who answered 'no' were classified as *Never-Tried*. Participants who responded that they have smoked more than five cigarettes to the question 'How many cigarettes have you smoked in your lifetime' were classified as *Smokers*; those who responded that they smoked every week or more were classified as *Weekly-Smokers*.

Please note that the study website contains details of all the data that is available through a fully searchable data dictionary (www.bris.ac.uk/alspac/researchers/data-access/data-dictionary). Ethical approval for the study was obtained from the ALSPAC ethics and Law Committee and the Local Research Ethics Committees.

IMAGEN: Smoking use was assessed using the European School Survey Project on Alcohol and Drugs questionnaires (1). Participants were asked 'On how many occasions during your lifetime have you smoked cigarettes?'. Seven options were given:
'0','1-2','3-5','6-9','10-19', '20-39' and '40 or more'.

Participants who responded zero were classified as *Never-Tried* and all the other as *Ever-Tried*. Participants who responded they have smoked more than five cigarettes were classified as *Smokers*. Participants were also asked how often they smoked and participants who were smoking every week or more frequently were classified as *Weekly-Smokers*. Although smoking behavior was assessed with different instruments in the four cohorts, the four categories were defined as similarly as possible.

Cotinine

Approximately 80% of nicotine is metabolized to cotinine in the liver by CYP2A6 enzyme (2). Cotinine is a relatively stable compound whose level reflects the cumulative intake of nicotine in the last week (3). Cotinine was measured only in the ALSPAC cohort from plasma samples collected at 15 years using the Cozart Cotinine Enzyme Immunoassay. Data was taken from the ALSPAC from EDTA blood plasma samples taken in a clinic assessment at 15 years. The plasma samples were stored at -80 °C and allowed to thaw at room temperature before use. Cotinine was measured using the Cozart Cotinine Enzyme Immunoassay (EIA) serum kit (M155B1). All samples were run in duplicates. Absorbance was measured spectrophotometrically at a wave length of 450 nm. Cotinine concentrations are expressed as ng/ml of blood. Cotinine level was quintile-transformed to reach normality. Quintile transformation has the advantage of including individuals with cotinine level equal to zero such as non-smokers in the analyses (4). Analyses using log-transformation (that exclude non-smokers) were also performed and showed substantially identical results (data not shown). Cotinine level was available in 2540 participants who also had genetic data available and were of European ancestry. Linear regression was performed to investigate the association between plasma continine level and smoking (see Supplementary-Table-3).

Genetic-association analyses within each sample

NFBC1966: Genome-wide genotyping was performed using the Illumina Infinium 370cnvDuo array. Quality control procedures used are described elsewhere (5). 33 directly genotyped SNPs covering the *TTC12-ANKK1-DRD2* were available.

NFBC1986: No genome-wide SNP data were available in this cohort. SNP rs2236709 was already genotyped as part of the custom Illumina Metabochip array (6). 13 additional SNPs were genotyped by Kbioscience (Hoddesdon, UK, http://www.kbioscience.co.uk/) using their

own system of fluorescence-based competitive allele-specific PCR (KASPar). SNPs were selected using a tagging approach to capture 75% of alleles at r^2>=0.8.

ALSPAC: Genome-wide genotyping was performed using the Illumina 550K array (Illumina, San Diego, CA, USA). We extracted the same SNPs available in NFBC1966.

IMAGEN: Participants were genotyped using the Illumina Quad 610 and 660 arrays. Among the SNPs that passed quality controls, 29 of the 33 SNPs were available.

Functional Magnetic Resonance Imaging (fMRI)

Monetary incentive delay (MID) task (7): Each trial involved an anticipation phase (4 sec), a response phase (2 sec), a feedback phase (2 sec), and a fixation period (4 sec). During the anticipation phase, cues indicating the amount of reward that could be won in a given trial (large, small, or none) were shown for 4 seconds. The participant could win large or small numbers of points (10 or 2) by responding as quickly as possible to a response cue. The points were converted to food snacks (small chocolate candies) following testing (5 points per candy). The participant completed 22 trials per condition, yielding 66 trials in total. One cue (circle with two lines) signaled that a large reward could be won, another (circle with one line) that a small reward could be won, and a third cue (triangle) that no reward could be won in the respective trial. Following a random time interval, a response cue was displayed, and the participant was instructed to respond as quickly as possible to this cue by means of a button press. The time window in which responses were counted as "wins" was adjusted dynamically during the course of the experiment according to the participant's performance, such that on average the participant won in 66% of all trials. The response and feedback phases had a total duration of 2 seconds. Four seconds of inter-trial fixation periods separated the trials.

Data Acquisition: Structural and functional Magnetic Resonance Imaging (fMRI) data were acquired by using 3-T MRI scanners from a range of manufacturers (Siemens, Philips, General Electric, Bruker), and the scanning variables were specifically chosen to be compatible with all scanners (8). For the present task, 300 volumes were acquired for each participant. Each volume consisted of 40 slices aligned to the line connecting the anterior-posterior commissure (2.4-mm thickness, 1-mm gap TR=2.20 s, TE=30 ms). The total scanning duration for the task took 11 min.

fMRI Preprocessing and Analysis

Structural and fMRI data were acquired using 3-T MRI scanners of different manufacturers (Siemens, GE, Philips). Pre-processing and data analyses were conducted using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) running in MATLAB® (9).

First, echo-planar images were co-registered with the T1 structural image. Functional images were realigned and resliced to the first volume. We created a custom template from the T_1 images of 552 adolescents by using the DARTEL toolbox (10) as implemented in SPM8. Single-subject contrast images were normalized to Montreal Neurological Institute (MNI) space by means of this custom template on the basis of the individual participants' DARTEL flow fields, and they were smoothed with a 5-mm Gaussian isotropic kernel. A first-level model was constructed on the unsmoothed single-participant data by using the following regressors: 1) anticipation of large reward, 2) anticipation of small reward, 3) anticipation of no reward, 4) feedback indicating large reward, 5) feedback indicating small reward, 6) feedback indicating no reward. Each regressor was included twice, once for successful trials, i.e., hits, and once for unsuccessful trials, i.e., misses. Trials in which participants failed to respond were modeled as separate error trials. Estimated movement

parameters were added to the design matrix in the form of 18 additional columns (3 translations, 3 rotations, 3 quadratic and 3 cubic translations, and each 3 translations with a shift of ±1 TR).

At the first level of analysis, changes in the blood-oxygen-level dependent (BOLD) response for each participant were assessed by linear combinations at the individual participant level, for each experimental condition; each trial (i.e. reward anticipation high reward) was convolved with the hemodynamic response function to form regressors that account for variance associated with the processing of reward anticipation and feedback. Contrast images of the parameter estimates were created for each participant.

The normalized and smoothed single-subject contrast images were then taken to a second-level random effects analysis to identify brain regions activating by anticipation of reward. Gender, site of recruitment and handedness were used as covariates of no interest. For the present investigation, we focused on the analyses on the reward anticipation phase using the contrast 'anticipation of large reward > anticipation of no reward'.

Region of interest (ROI) analyses were run to investigate the association between activations in the main site of reward processing, i.e. the ventral striatum, and genotypes. The ROIs for the bilateral ventral striatum were extracted based on coordinates from previous findings (11) from the contrast 'anticipation of high reward > anticipation of no reward' as 9-mm sphere (x= ±15, y=9, z=-9) using MarsBaR version 0.42 (12), see Supplementary-Figure-2.

From all participants who provided fMRI data for the MID task (n=1668), 1610 and 1630 for the left and right ventral striatum, respectively, were left after outlier removal according to the following criterion: Mean +/- 2*SD. 1263 of these had genotypic data as well and these were included in the analyses between fMRI and genotypic data.

Genetic-association analyses within each sample:

To control for population stratification, principal components were computed from genome-wide data (NFBC1966, ALSPAC and IMAGEN) (13) or Multidimensional Scaling coordinates (14) computed using 130,000 SNPs from the Illumina metabochip array (NFBC1986) (6) and included as covariates into the analysis.

Statistical Analyses

Meta-analysis: The meta-analysis procedure in Plink v1.07 (15) requires input files with odds ratio, beta estimates, confidence interval, and p-value for each study. Meta-analysis outputs include a beta, odds ratios, and a p-value for fixed and random effect models. The fixed effects model assumes that the effect of the independent variable on the outcome is the same in each study. The random effect model allows the effect to vary between studies. The output gives two measures: Cochrane's *Q* statistic and I^2 heterogeneity index. The I^2 statistic lies between 0%-100% with I^2=0 indicating no heterogeneity (16).

TABLES

Supplementary Table 1. Minor allele frequencies (MAF) in the 4 different cohorts.

BP	Gene	SNP	Location	MAF					Ref allele/	N	
				Overall	Imagen	Alspac	1966	1986	Other allele	Samples	Individuals
112685412	*TTC12*	rs4517559	flanking 5'UTR	0.3736	0.3681	0.3586	0.4920	0.4944	C/T	4	8565
112691986	*TTC12*	rs2236709	Intron	0.2588	0.2533	0.2396	0.2961	0.3031	G/A	4	8866
112693407	*TTC12*	rs7927508	Intron	0.3667	0.3637	0.3578	0.4920	NA	G/A	3	6336
112694133	*TTC12*	rs2156486	Intron	0.1814	0.1817	0.2574	0.1865	0.1801	G/T	4	8569
112699378	*TTC12*	rs723077	Coding, Met73Leu	0.4894	0.4955	0.4977	0.4260	*NA*	C/A	3	6344
112704356	*TTC12*	rs10502172	Intron	0.4596	0.4685	0.4851	0.4616	0.4526	T/C	4	8633
112705919	*TTC12*	rs2303380	Intron splice site	0.3819	0.3824	0.3664	0.2970	*NA*	G/A	3	6344
112716539	*TTC12*	rs2288159	Intron	0.1561	0.1480	0.1479	0.2417	*NA*	T/G	3	6343
112721133	*TTC12*	rs4987094	Intron	0.1004	0.0977	0.0995	0.1979	*NA*	A/G	3	6347
112735810	*TTC12*	rs2276070	Intron	0.1541	0.1481	0.1478	0.2410	*NA*	T/C	3	6345
112739889	*TTC12*	rs719802	Intron	0.3953	0.3915	0.3970	0.3242	*NA*	T/C	3	6344
112739985	*TTC12*	rs719804	Intron	0.249	0.2500	0.1941	0.1732	*NA*	G/A	3	6334

112749387	*TTC12*	rs2282511	flanking 3'UTR	0.3455	*NA*	0.3274	0.2957	0.2959	A/C	3	7145
112754346	*TTC12*	rs754672	flanking 3'UTR	0.4894	0.4831	0.4621	0.4629	0.4513	T/C	4	8552
112761718	*ANKK1*	rs877138	flanking 5'UTR	0.3431	0.3456	0.3336	0.3003	*NA*	G/A	3	6347
112768580	*ANKK1*	rs4590907	Intron	0.1436	0.1397	0.1362	0.2342	0.2487	G/T	4	8578
112772031	*ANKK1*	rs7118900	Coding, Ala239Thr	0.1896	0.1870	0.1869	0.1565	*NA*	A/G	3	6339
112775225	*ANKK1*	rs4938016	Coding, Gly442Arg	0.44529	*NA*	0.3059	0.3775	0.0000	G/C	3	7239
112775370	*ANKK1*	rs2734849	Coding, His490Arg	0.4886	0.4976	0.4925	0.4624	0.4750	G/A	4	8575
112776038	*ANKK1*	rs1800497	Coding, Glu713Lys,TaqI A	0.2026	0.2015	0.1976	0.1703	0.1806	A/G	4	8561
112788669	*DRD2*	rs6277	Coding, Pro219Pro	0.12201	*NA*	0.4505	0.4658	0.4537	A/G	3	7249
112801119	*DRD2*	rs1076563	Intron	0.4127	0.4041	0.3924	0.4972	*NA*	C/A	3	6346
112815891	*DRD2*	rs7125415	Intron	0.09914	0.0944	0.0959	0.1896	*NA*	T/C	3	6346
112818599	*DRD2*	rs4648318	Intron	0.2538	0.2473	0.2439	0.3407	*NA*	C/T	3	6346
112824662	*DRD2*	rs4274224	Intron	0.4823	0.4814	0.4917	0.2391	*NA*	G/A	3	6342
112829684	*DRD2*	rs4581480	Intron	0.09919	0.0954	0.1027	0.0683	*NA*	C/T	3	6344
112834984	*DRD2*	rs7131056	Intron	0.4205	0.4202	0.4264	0.4942	*NA*	C/A	3	6341
112846601	*DRD2*	rs4938019	Intron	0.1447	0.1441	0.1401	0.2335	0.2392	C/T	4	6334

112852165	*DRD2*	rs12364283	flanking 5'UTR	0.07563	0.0779	0.0775	0.0787	*NA*	G/A	3	8583
112857971	*DRD2*	rs10891556	intergenic	0.1758	0.1751	0.1734	0.2380	*NA*	T/G	3	6339
112860946	*DRD2*	rs6589377	intergenic	0.3776	*NA*	0.3820	0.1689	*NA*	G/A	4	6345
112863421	*DRD2*	rs4482060	intergenic	0.30032	*NA*	0.4330	0.2589	0.2545	T/A	3	8576

Supplementary Table 2. Association between SNPs spanning the *TTC12-ANKK1-DRD2* gene-cluster and self-reported smoking behavior for *Smokers* vs *Never-Tried*.

Notes: *Smokers* (N=2819) are compared *Never-Tried* (N=6049) (binary logistic regression). Results from NFBC1966, NFBC1986, IMAGEN and ALSPAC are combined using meta-analysis. FEM=Fixed effect model, REM=Random effect model, *Q*=Cochrane's *Q* statistic, I^2=heterogeneity index (0-100). Significant p values are in bold, of these those that remain significant after Bonferroni correction for multiple testing are indicated by an asterisk. Thick lines indicate LD blocks.

BP	Gene	SNP	Location	MAF	Ref allele/ Other allele	N		*P*		*OR*		*Q*	I^2
						Samples	Individuals	(FEM)	(REM)	(FEM)	(REM)		
112685412	*TTC12*	rs4517559	flanking 5'UTR	0.3736	C/T	4	8565	0.01		1.09		0.26	25
112691986	*TTC12*	rs2236709	Intron	0.2588	G/A	4	8866	0.001*		1.13		0.46	0
112693407	*TTC12*	rs7927508	Intron	0.3667	G/A	3	6336	0.008		1.12		0.26	25.84
112694133	*TTC12*	rs2156486	Intron	0.1814	G/T	4	8569	0.022		0.9		0.46	0
112699378	*TTC12*	rs723077	Coding, Met73Leu	0.4894	C/A	3	6344	0.3		0.96		0.39	0
112704356	*TTC12*	rs10502172	Intron	0.4596	T/C	4	8633		0.14		0.9	0.01	76.24
112705919	*TTC12*	rs2303380	Intron splice site	0.3819	G/A	3	6344		0.83		1.02	0.001	84.44
112716539	*TTC12*	rs2288159	Intron	0.1561	T/G	3	6343	0.009		1.14		0.86	0
112721133	*TTC12*	rs4987094	Intron	0.1004	A/G	3	6347	0.013		1.16		0.94	0

112735810	*TTC12*	rs2276070	Intron	0.1541	T/C	3	6345	0.009		1.16		0.81	0
112739889	*TTC12*	rs719802	Intron	0.3953	T/C	3	6344	0.21	0.66	1.05	1.04	0.011	77.98
112739985	*TTC12*	rs719804	Intron	0.249	G/A	3	6334	0.09		1.09		0.19	39.05
112749387	*TTC12*	rs2282511	flanking 3'UTR	0.3455	A/C	3	7145		0.19		1.11	0.01	77.25
112754346	*TTC12*	rs754672	flanking 3'UTR	0.4894	T/C	4	8552		0.07		0.89	0.02	68.26
112761718	*ANKK1*	rs877138	flanking 5'UTR	0.3431	G/A	3	6347		0.65		1.05	0.004	81.87
112768580	*ANKK1*	rs4590907	Intron	0.1436	G/T	4	8578	0.05		1.08		0.43	0
112772031	*ANKK1*	rs7118900	Coding, Ala239Thr	0.1896	A/G	3	6339	0.78		0.99		0.58	0
112775225	*ANKK1*	rs4938016	Coding, Gly442Arg	0.44529	G/C	3	7239		0.06		1.12	0.08	61.34
112775370	*ANKK1*	rs2734849	Coding, His490Arg	0.4886	G/A	4	8575		0.09		0.9	0.03	66.72
112776038	*ANKK1*	rs1800497	Coding, Glu713Lys,TaqI	0.2026	A/G	4	8561	0.85		1.01		0.72	0

			A										
112788669	*DRD2*	rs6277	Coding, Pro219Pro	0.12201	A/G	3	7249		0.03		0.89	0.09	59.03
112801119	*DRD2*	rs1076563	Intron	0.4127	C/A	3	6346	0.003		0.89		0.14	48.44
112815891	*DRD2*	rs7125415	Intron	0.09914	T/C	3	6346	0.01		1.16		0.77	0
112818599	*DRD2*	rs4648318	Intron	0.2538	C/T	3	6346	0.003		1.14		0.18	41.98
112824662	*DRD2*	rs4274224	Intron	0.4823	G/A	3	6342	0.93		1		0.14	49.82
112829684	*DRD2*	rs4581480	Intron	0.09919	C/T	3	6344		0.53		1.07	0.1	56.29
112834984	*DRD2*	rs7131056	Intron	0.4205	C/A	3	6341	0.22		1.05		0.31	15.05
112846601	*DRD2*	rs4938019	Intron	0.1447	C/T	4	6334	0.09		1.08		0.52	0
112852165	*DRD2*	rs12364283	flanking 5'UTR	0.07563	G/A	3	8583	0.88		1.01		0.88	0
112857971	*DRD2*	rs10891556	intergenic	0.1758	T/G	3	6339	0.51		1.03		0.38	0
112860946	*DRD2*	rs6589377	intergenic	0.3776	G/A	4	6345		0.86		1.01	0.06	59.84
112863421	*DRD2*	rs4482060	intergenic	0.30032	T/A	3	8576	0.68		1.02		0.14	48.38

Supplementary Table 3. Association between SNPs spanning the *TTC12-ANKK1-DRD2* gene-cluster and self-reported smoking behavior for *Weekly-Smokers* vs *Never-Tried*.

Notes: *Weekly-Smokers* (N=1624) are compared to *Never-Tried* (N=6049) (binary logistic regression). Results from NFBC1966, ALSPAC, NFBC1986, and IMAGEN are combined using meta-analysis. FEM=Fixed effect model, REM=Random effect model, *Q*=Cochrane's *Q* statistic, I^2=heterogeneity index (0-100). Significant p values are in bold (uncorrected, Bonferroni corrected threshold: 0.0015 are indicated by *). Thick lines indicate LD blocks.

BP	**Gene**	**SNP**	**Location**	**MAF**	**Ref/**	**N**		***P***		***OR***		***Q***	**I^2**
					Other allele	**Samples**	**Individuals**	**(FEM)**	**(REM)**	**(FEM)**	**(REM)**		
112685412	*TTC12*	rs4517559	flanking 5'UTR	0.3736	C/T	4	7385	**0.04**		1.09		0.72	0
112691986	*TTC12*	rs2236709	Intron	0.2588	G/A	4	7672	**0.001***		1.16		0.94	0
112693407	*TTC12*	rs7927508	Intron	0.3667	G/A	3	5281	**0.05**		1.12		0.62	0
112694133	*TTC12*	rs2156486	Intron	0.1814	G/T	4	7390	**0.04**		0.89		0.61	0

112699378	*TTC12*	rs723077	Coding, Met73Leu	0.4894	C/A	3	5287	0.45		0.96		0.43	0
112704356	*TTC12*	rs10502172	Intron	0.4596	T/C	4	7456		0.08		0.89	0.08	56.11
112705919	*TTC12*	rs2303380	Intron splice site	0.3819	G/A	3	5288		0.63		1.06	0.03	72.38
112716539	*TTC12*	rs2288159	Intron	0.1561	T/G	3	5286	0.11		1.12		0.96	0
112721133	*TTC12*	rs4987094	Intron	0.1004	A/G	3	5290	0.09		1.15		0.8	0
112735810	*TTC12*	rs2276070	Intron	0.1541	T/C	3	5289	0.11		1.12		0.97	0
112739889	*TTC12*	rs719802	Intron	0.3953	T/C	3	5289		0.43		1.07	0.12	52.08
112739985	*TTC12*	rs719804	Intron	0.249	G/A	3	5278		0.22		1.14	0.12	53.62
112749387	*TTC12*	rs2282511	flanking 3’UTR	0.3455	A/C	3	6077		0.07		1.14	0.12	52.19
112754346	*TTC12*	rs754672	flanking 3’UTR	0.4894	T/C	4	7377	**0.008**		0.89		0.13	46.17
112761718	*ANKK1*	rs877138	flanking 5’UTR	0.3431	G/A	3	5290		0.46		1.08	0.08	60.56

112768580	*ANKK1*	rs4590907	Intron	0.1436	G/T	4	7398	0.41		1.04		0.79	0
112772031	*ANKK1*	rs7118900	Coding, Ala239Thr	0.1896	A/G	3	5283	0.84		0.99		0.7	0
112775225	*ANKK1*	rs4938016	Coding, Gly442Arg	0.44529	G/C	3	6153	**0.01**		1.12		0.14	48.79
112775370	*ANKK1*	rs2734849	Coding, His490Arg	0.4886	G/A	4	7397	**0.005**		0.89		0.34	10.17
112776038	*ANKK1*	rs1800497	Coding, Glu713Lys	0.2026	A/G	4	7386	0.65		1.02		0.82	0
112788669	*DRD2*	rs6277	Coding, Pro219Pro	0.12201	A/G	3	6161	**0.005**		0.88		0.39	0
112801119	*DRD2*	rs1076563	Intron	0.4127	C/A	3	5289	**0.03**		0.88		0.55	0
112803549	*DRD2*	rs2471857	Intron	0.1577	T/C	3	5289	0.91		1.01		0.54	0
112815891	*DRD2*	rs7125415	Intron	0.09914	T/C	3	5289	0.24		1.1		0.91	0
112818599	*DRD2*	rs4648318	Intron	0.2538	C/T	3	5286	**0.02**		1.16		0.32	11.96
112824662	*DRD2*	rs4274224	Intron	0.4823	G/A	3	5287	0.85		0.99		0.45	0

112829684	*DRD2*	rs4581480	Intron	0.09919	C/T	3	5284		0.87		1.03	0.09	57.52
112834984	*DRD2*	rs7131056	Intron	0.4205	C/A	3	5280	0.09		1.1		0.61	0
112846601	*DRD2*	rs4938019	Intron	0.1447	C/T	4	7404	0.06		1.1		0.99	0
112852165	*DRD2*	rs12364283	flanking 5'UTR	0.07563	G/A	3	5283	0.63		1.05		0.22	34.83
112857971	*DRD2*	rs10891556	intergenic	0.1758	T/G	3	5289	0.26		1.09		0.83	0
112860946	*DRD2*	rs6589377	intergenic	0.3776	G/A	4	7396	0.58		1.03		0.51	0
112863421	*DRD2*	rs4482060	intergenic	0.30032	T/A	3	6148	0.18		1.07		0.466	0

Supplementary Table 4. Association between SNPs spanning the *TTC12-ANKK1-DRD2* gene-cluster and cotinine level (quantile-transformed in the ALSPAC cohort (N=2,540) (linear regression). Significant p values are in bold (uncorrected, Bonferroni corrected threshold: 0.0015 are indicated by *). Thick lines indicate LD blocks.

SNP	**Gene**	**BP**	**Location**	**MAF**	***Ref allele/ Other allele***	**BETA**	**STAT**	***p***
rs4517559	*TTC12*	1.1E+08	flanking 5'UTR	0.3736	C/T	0.09	3.18	**0.001***
rs2236709	*TTC12*	1.1E+08	Intron	0.2588	G/A	0.11	3.49	**0.0005***
rs7927508	*TTC12*	1.1E+08	Intron	0.3667	G/A	0.09	3.16	**0.002**
rs2156486	*TTC12*	1.1E+08	Intron	0.1814	G/T	0.03	0.82	0.41
rs723077	*TTC12*	1.1E+08	Coding, Met73Leu	0.4894	C/A	-0.03	-1.2	0.23
rs10502172	*TTC12*	1.1E+08	Intron	0.4596	T/C	-0.01	-0.36	0.72
rs2303380	*TTC12*	1.1E+08	Intron splice site	0.3819	G/A	-0.02	-0.77	0.49
rs2288159	*TTC12*	1.1E+08	Intron	0.1561	T/G	0.06	1.58	0.11
rs4987094	*TTC12*	1.1E+08	Intron	0.1004	A/G	0.04	0.82	0.41
rs2276070	*TTC12*	1.1E+08	Intron	0.1541	T/C	0.06	1.57	0.12
rs719802	*TTC12*	1.1E+08	Intron	0.3953	T/C	-0.01	-0.29	0.78
rs719804	*TTC12*	1.1E+08	Intron	0.249	G/A	-0.01	-0.4	0.69

rs2282511	*TTC12*	1.1E+08	flanking 3'UTR	0.3455	A/C	0	0.15	0.88
rs754672	*TTC12*	1.1E+08	flanking 3'UTR	0.4894	T/C	-0.03	-1.19	0.23
rs877138	*ANKK1*	1.1E+08	flanking 5'UTR	0.3431	G/A	0.01	0.35	0.73
rs4590907	*ANKK1*	1.1E+08	Intron	0.1436	G/T	0.07	1.7	0.09
rs7118900	*ANKK1*	1.1E+08	Coding, Ala239Thr	0.1896	A/G	-0.05	-1.36	0.18
rs4938016	*ANKK1*	1.1E+08	Coding, Gly442Arg	0.44529	G/C	0.07	2.34	**0.02**
rs2734849	*ANKK1*	1.1E+08	Coding, His490Arg	0.4886	G/A	-0.03	-1.17	0.24
rs1800497	*ANKK1*	1.1E+08	Coding, Glu713Lys, TaqI A	0.2026	A/G	-0.04	-0.98	0.33
rs6277	*DRD2*	1.1E+08	Coding, Pro219Pro	0.12201	A/G	-0.05	-1.81	0.07
rs1076563	*DRD2*	1.1E+08	Intron	0.4127	C/A	-0.05	-1.64	0.1
rs2471857	*DRD2*	1.1E+08	Intron	0.1577	T/C	-0.02	-0.42	0.67
rs7125415	*DRD2*	1.1E+08	Intron	0.09914	T/C	0.07	1.45	0.15
rs4648318	*DRD2*	1.1E+08	Intron	0.2538	C/T	0.06	1.97	**0.05**
rs4274224	*DRD2*	1.1E+08	Intron	0.4823	G/A	-0.03	-1.03	0.3
rs4581480	*DRD2*	1.1E+08	Intron	0.09919	C/T	-0.01	-0.2	0.84

rs7131056	*DRD2*	1.1E+08	Intron	0.4205	C/A	0	-0.07	0.94
rs4938019	*DRD2*	1.1E+08	Intron	0.1447	C/T	0.05	1.36	0.17
rs12364283	*DRD2*	1.1E+08	flanking 5'UTR	0.07563	G/A	0.05	0.87	0.38
rs10891556	*DRD2*	1.1E+08	intergenic	0.1758	T/G	0.04	1.06	0.29
rs6589377	*DRD2*	1.1E+08	intergenic	0.3776	G/A	-0.04	-1.24	0.22
rs4482060	*DRD2*	1.1E+08	intergenic	0.30032	T/A	-0.03	-1.01	0.31

FIGURE LEGENDS:

Supplementary-Figure 1: Linkage disequilibrium (LD) of the chromosome 11q23 region in the NFBC1966 **(A)**, ALSPAC **(B)**, IMAGEN **(C)** and NFBC1986 **(D)** cohorts. Three haplotype blocks are indicated with borders. LD is coded according to the following color scheme: LOD < 2 and D' < 1: white; LOD < 2 and D' = 1: blue; LOD ≥ 2 and D' < 1: shades of red; LOD ≥ 2 and D' = 1: bright red. The *TTC12-ANKK1-DRD2* gene-cluster covers a large genomic area (chromosome 11 *TTC12*: 112,690,539-112,749,226, 58,688 bp; *ANKK1*: 112,763,723-112,776,350, 12,628 bp; *DRD2* short isoform: 112,785,528-112,851,103, 12,628 bp and *DRD2* long isoform: 112,785,528-112,851,103, 65,576 bp).

Supplementary-Figure 2: (A) Box plot of blood-oxygen level-dependent (BOLD) response in left ventral striatum stratified by rs2236709 genotype (note that G-carriers are pooled together). The x-axis represents the number of minor G-alleles (N_{AA}=706, $N_{AG/GG}$=557).

Mean BOLD response within each group was as follows: M_{AA}=0.365, SD_{AA}=0.370 and $M_{AG/GG}$=0.417, $SD_{AG/GG}$=0.391. (**B)** Region of interest as extracted from the contrast 'anticipation of high reward > anticipation for no reward' from the modified Monetary Incentive Delay (MID) task in the IMAGEN sample (N=1263) indicating left and right ventral striatum centred at Montreal Neurological Institute (MNI): x=±15, y=9, z=-9.

REFERENCES

1. Hibell B AB, Bjarnason T, Ahlstro¨m S, Balakireva O,, A K. The ESPAD Report: Alcohol and
Other Drug Use Among Students in 35 European Countries. Modintryckoffset AB: Stockholm, 2003.

2. Hukkanen J, Jacob P, 3rd, Benowitz NL. Metabolism and disposition kinetics of nicotine. Pharmacol Rev. 2005; 57(1): 79-115.

3. Benowitz NL, Kuyt F, Jacob P, 3rd, Jones RT, Osman AL. Cotinine disposition and effects. Clinical Pharmacology & Therapeutics. 1983; 34(5): 604-11.

4. Schumann G, Coin LJ, Lourdusamy A, Charoen P, Berger KH, Stacey D, et al. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. Proc Natl Acad Sci U S A. 2011; 108(17): 7119-24.

5. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nature genetics. 2009; 41(1): 35-46.

6. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nature genetics. 2012.

7. Peters J, Bromberg U, Schneider S, Brassen S, Menz M, Banaschewski T, et al. Lower ventral striatal activation during reward anticipation in adolescent smokers. The American journal of psychiatry. 2011; 168(5): 540-9.

8. Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Buchel C, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Molecular psychiatry. 2010; 15(12): 1128-39.

9. The MathWorks Inc. MATLAB. The MathWorks Inc., 2011.

10. Ashburner J. A fast diffeomorphic image registration algorithm. Neuroimage. 2007; 38(1): 95-113.

11. Yacubian J, Glaescher J, Schroeder K, Sommer T, Braus DF, Buechel C. Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. Journal of Neuroscience. 2006; 26(37): 9530-7.

12. Brett M, Anton J-L, Valabregue R, Poline J-B. Region of interest analysis using the MarsBar toolbox for SPM 99. Neuroimage. 2002; 16: S497.

13. Ducci F, Enoch MA, Yuan Q, Shen PH, White KV, Hodgkinson C, et al. HTR3B is associated with alcoholism with antisocial behavior and alpha EEG power--an intermediate phenotype for alcoholism and co-morbid behaviors. Alcohol. 2009; 43(1): 73-84.

14. Wang D, Sun Y, Stang P, Berlin JA, Wilcox MA, Li Q. Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. BMC Proc. 2009; 3 Suppl 7: S109.

15. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007; 81(3): 559-75.

16. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. British Medical Journal. 2003; 327(7414): 557-60.